Problems of Information Transmission,
vol. 50, no. 3, pp. 19–34, 2014.

M. V. Burnashev[1], H. Yamamoto[2]

ON USING FEEDBACK IN A GAUSSIAN CHANNEL

For information transmission a discrete time channel with independent additive Gaussian noise is used. There is also another channel with independent additive Gaussian noise (the feedback channel), and the transmitter observes without delay all outputs of the forward channel via that channel. Transmission of nonexponential number of messages is considered (i.e. transmission rate equals zero) and the achievable decoding error exponent for such a combination of channels is investigated. The transmission method strengthens the method used by authors earlier for BSC and Gaussian channels. In particular, for small feedback noise, it allows to gain 33.3% (instead of 23.6% earlier in the similar case of Gaussian channel).

§ 1. Introduction and main result

In the paper results of [1] are strengthened and proofs are simplified. We consider the discrete time channel with independent additive Gaussian noise, i.e. if $\boldsymbol{x} = (x_1, \dots, x_n)$ is the input codeword then the received block $\boldsymbol{y} = (y_1, \dots, y_n)$ is

$$y_i = x_i + \xi_i, \qquad i = 1, \dots, n, \tag{1}$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ are independent $\mathcal{N}(0,1)$–Gaussian random variables, i.e. $\mathbf{E}\xi_i = 0$, $\mathbf{E}\xi_i^2 = 1$. There is the noisy feedback channel, and the transmitter observes (without delay) all outputs $\{z_i\}$ of the forward channel via that noisy feedback channel

$$z_i = y_i + \sigma\eta_i, \qquad i = 1, \dots, n, \tag{2}$$

where $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$ are independent (and independent of ξ) $\mathcal{N}(0,1)$–Gaussian random variables, i.e. $\mathbf{E}\eta_i = 0$, $\mathbf{E}\eta_i^2 = 1$. The value $\sigma > 0$, characterizing feedback channel noise intensity, is given. No coding is used in the feedback channel (i.e. the receiver simply retransmits all received outputs to the transmitter). In other words, the feedback channel is "passive".

[1]Supported in part by the Russian Foundation for Basic Research, project nos. 12-01-00905a and 13-01-12458 ofi_m2.

[2]Supported in part by the Japanese Fund of JSPS KAKENHI, grant no. 25289111.

We assume that the input block $\boldsymbol{x}$ satisfies the constraint

$$\sum_{i=1}^{n} x_i^2 \le nA, \tag{3}$$

where A is a given constant. We denote by AWGN(A) the channel (1) with constraint (3) without feedback, and by AWGN(A, σ) that channel with noisy feedback (2). The capacity of both channels equals $C(A) = [\ln(1 + A)]/2$.

We consider the case when the overall transmission time n and $M = e^{o(n)}$, $n \to \infty$, equiprobable messages $\{\theta_1, \ldots, \theta_M\}$ are given. After the moment n, the receiver makes a decision $\hat{\theta}$ on the message transmitted. We are interested in the best possible decoding error exponent (and whether it exceeds the similar exponent of the channel without feedback).

It is well known [2] that even noiseless feedback does not increase the capacity of the Gaussian channel (or any other memoryless channel). However, feedback allows to improve the decoding error exponent (*channel reliability function*) with respect to no-feedback channel. Possibility of such improvement stimulated a good interest to that topic in 60–80's. A good number of interesting results have been obtained during that period (e.g. [3–10]). Unfortunately, all those papers had a common drawback: their methods were heavily based on the assumption that the feedback is *noiseless*. It was necessary in order to have perfect mutual coordination between both the transmitter and the receiver. Essentially, any noise in the feedback link destroyed that coordination and all hypothetical improvements. It was not clear whether it is possible to improve communication characteristics using more realistic noisy feedback.

That uncertainty with noisy feedback remained till 2008, when in [11]–[14] it was shown (for BSC) how to use such feedback in order to improve the decoding error exponent. Although improvement was not large (approximately 14.3%. for small feedback noise), it was the first method that worked for noisy feedback. Later results ([1] and this paper) are developments of [11]–[14].

In order to explain what is new in the paper, remind briefly what was done in earlier papers [11]–[14] and [1]. For that purpose we explain first why noiseless feedback allows to improve decoding error exponent. For a channel without feedback that exponent is determined (for small transmission rates R) by the code distance of the code used (i.e. by the minimal distance among codewords). Noiseless feedback allows during transmission to change the code (code function) used, e.g. increasing the distances among most probable codewords. That feature allowed to improve the decoding error exponent. But for that purpose an ideal coordination between both the transmitter and the receiver are required.

In all papers [11]–[14], [1] and this one coding function can be changed only at one fixed moment ("switching moment"). In [11]–[14] such change took place only if two most probable codewords were much more probable than all remaining codewords. It was shown that if noise in the feedback channel is less than a certain critical value p_{crit}, then it is possible to choose transmission parameters such that the probability of miscoordination between the transmitter and the receiver becomes smaller than decoding error probability. That fact allowed to improve the decoding error exponent with respect to no-feedback channel.

Later in the paper [1] for Gaussian channel that method was strengthened taking into account not two, but three most probable codewords. Moreover, the decoding method was improved. It allowed not only to improve the gain (23.6% instead of 14.3% in [12], but also to show that for any *noise intensity* $\sigma^2 < \infty$ it is possible to improve the best error exponent of AWGN(A) no-feedback channel. Of course, if σ is not small then the gain is small, but it is strongly positive. In other words, in the problem considered there is no any critical level $\sigma_{\rm crit}$, beyond which it is not possible to improve the error exponent of the no-feedback channel. It should be noticed also that the investigation method with optimal decoding in [1] was rather tedious.

The method of papers [11, 12] was applied to Gaussian channel AWGN(A, σ) in [15] with similar to [11, 12] results (in particular, with the same asymptotic gain 14.3%).

The aim of the paper is to strengthen the transmission method [1] (in particular, using up to four most probable codewords) and also simplify its analysis. It allows to improve the gain up to 33.3% (instead of 23.6% in [1]).

Remark 1. We consider the case when the value $\sigma^2 > 0$ is fixed and does not depend on the number of messages M.

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ denote

$$(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} x_i y_i, \quad \|\boldsymbol{x}\|^2 = (\boldsymbol{x}, \boldsymbol{x}), \quad d\,(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2.$$

A subset $\mathcal{C} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ with $\|\boldsymbol{x}_i\|^2 = An$, $i = 1, \ldots, M$ is called a (M, A, n)–code of length n.

For a code $\mathcal{C} = \{\boldsymbol{x}_i\}$ denote by $P_{\rm e}(\mathcal{C})$ the minimal possible decoding error probability

$$P_{\rm e}(\mathcal{C}) = \min \max_i P(e|\boldsymbol{x}_i),$$

where $P(e|\boldsymbol{x}_i)$ – conditional decoding error probability provided $\boldsymbol{x}_i$ was transmitted, and minimum is taken over all decoding methods (it will be convenient to denote the message transmitted as θ_i and $\boldsymbol{x}_i$ as well).

In the paper we consider the case when $M = M_n \to \infty$, but $M_n = e^{o(n)}$ as $n \to \infty$ (it corresponds to zero-rate of transmission). For M messages and AWGN(A) channel denote by $P_{\rm e}(M, A, n)$ the minimal possible decoding error probability for the best (M, A, n)–code and introduce the exponent (in n) of that function [16]

$$E(A) = \limsup_{\substack{n\to\infty \\ M\to\infty \\ \ln M = o(n)}} \frac{1}{n} \ln \frac{1}{P_{\rm e}(M, A, n)} = \frac{A}{4}. \tag{4}$$

Similarly, for AWGN(A, σ) channel with noisy feedback denote by $P_{\rm e}(M, A, \sigma, n)$ the minimal possible decoding error probability and introduce the function

$$F(A, \sigma) = \limsup_{\substack{n\to\infty \\ M\to\infty \\ \ln M = o(n)}} \frac{1}{n} \ln \frac{1}{P_{\rm e}(M, A, \sigma, n)}.$$

It is also known that if $\sigma = 0$ (i.e. noiseless feedback) then [7]

$$F(A, 0) = \frac{A}{2}. \tag{5}$$

For AWGN(A, σ) channel denote by $F_1(A, \sigma)$ the best error exponent for the transmission method with one switching moment, described in §2. Then $F_1(A, \sigma) \le F(A, \sigma)$ for all A, σ.

The paper main result is as follows.

T h e o r e m. *Let $M \to \infty$ and $\ln M = o(n)$, $n \to \infty$. Then the formula holds*

$$F_1(A, \sigma) \ge \frac{A(1 - \sigma^2)}{3}. \tag{6}$$

For small σ the formula (6) gives 33.3% of improvement with respect to no-feedback channel (see (4)). It is given in a simplified form oriented to small values of σ. A more general formula (following from results of §4) would be too bulky.

Remark 2. The method described in the paper and its analysis can be generalized on slow growing number $N = N(\sigma)$ of switches. It allows to prove the following result

$$F_{N(\sigma)}(A, \sigma) = \frac{A(1 + o(\sigma))}{2}, \qquad \sigma \to 0. \tag{7}$$

In other words, for small σ the formula (7) gives improvement of 100% with respect to no-feedback channel (see (4)), and it coincides with similar result (5) for noiseless feedback. It will be done in another paper.

In §2 the transmission method with one switching moment and its decoding are described. In §§ 3-4 its analysis is performed and the theorem is proved. Greek letters $\xi, \eta, \zeta, \xi_1, \dots$ in the paper designate $\mathcal{N}(0, 1)$–Gaussian random variables.

§ 2. Transmission/decoding method

We use the transmission strategy with one fixed switching moment at which the code used will be changed. Denote $n_1 = n/2$ and partition the total transmission time $[1, n]$ on two phases: $[1, n_1]$ (phase I) and $[n_1 + 1, n]$ (phase II). After moment n the receiver makes a decision in favor of the most probable message θ_i (based on all received on $[1, n]$ signals).

Each of M codewords $\{\boldsymbol{x}_i\}$ of length n have the form $\boldsymbol{x}_i = (\boldsymbol{x}'_i, \boldsymbol{x}''_i)$, where both $\boldsymbol{x}'_i$ (to be used on phase I) and $\boldsymbol{x}''_i$ (to be used on phase II) have length n_1 .

Similarly, the received block $\boldsymbol{y}$ has the form $\boldsymbol{y} = (\boldsymbol{y}', \boldsymbol{y}'')$, where $\boldsymbol{y}'$ is the block received on phase I and $\boldsymbol{y}''$ is the block received on phase II. Denote by $\boldsymbol{z}'$ the received (by the transmitter) block on phase I. The codewords first parts $\{\boldsymbol{x}'_i\}$ are fixed, while the second parts $\{\boldsymbol{x}''_i\}$ will depend on the block $\boldsymbol{z}'$ received by the transmitter on phase I.

We set two positive constants A_1, A_2 such that

$$A_1 + A_2 = nA, \tag{8}$$

and denote

$$\beta = \frac{A_2}{A_1}. \tag{9}$$

Then $A = (1+\beta)A_1/n$. At the end of Theorem proof we set $\beta = 1/2$.

Denoting

$$d_i = d(\boldsymbol{x}'_i, \boldsymbol{y}') = \|\boldsymbol{y}' - \boldsymbol{x}'_i\|^2,$$

arrange the distances $\{d_i, i = 1, \dots, M\}$ for the receiver after phase I in the increasing order, and denote

$$d^{(1)} = \min_i d_i \le d^{(2)} \le \ldots \le d^{(M)} = \max_i d_i$$

(case of tie has zero probability). Let also $\boldsymbol{x}'^{(1)}, \dots, \boldsymbol{x}'^{(M)}$ be the corresponding ranking of codewords $\{\boldsymbol{x}'\}$ after phase I for the receiver, i.e $\boldsymbol{x}'^{(1)}$ is the closest to $\boldsymbol{y}'$ codeword, etc.

Similarly, denoting

$$d_i^{(t)} = d(\boldsymbol{x}'_i, \boldsymbol{z}') = \|\boldsymbol{z}' - \boldsymbol{x}'_i\|^2,$$

arrange the distances $\{d_i^{(t)}, i = 1, \dots, M\}$ for the transmitter after phase I in the increasing order, denoting

$$d^{(1)t} = \min_i d_i^{(t)} \le d^{(2)t} \le \ldots \le d^{(M)t} = \max_i d_i^{(t)}.$$

Let also $\boldsymbol{x}'^{(1)t}, \dots, \boldsymbol{x}'^{(M)t}$ be the corresponding ranking of codewords $\{\boldsymbol{x}'\}$ after phase I for the transmitter, i.e $\boldsymbol{x}'^{(1)t}$ is the closest to $\boldsymbol{z}'$ codeword, etc.

Transmission method with one switching moment. We choose a set $\mathcal{K}$ of codes $\mathcal{C}$ which the transmitter may use on phase II. A code $\mathcal{C} \in \mathcal{K}$ used on phase II depends on the received block $\boldsymbol{z}'$. Based on $\boldsymbol{y}'$, the receiver finds the probability distribution $\mathbf{P}_r(\mathcal{C}|\boldsymbol{y}')$, $\mathcal{C} \in \mathcal{K}$ of the code $\mathcal{C}$ used by the transmitter on phase II, and uses that distribution for optimal decoding. It is a crucial point of the whole method.

Transmission. In order to simplify exposition it is sufficient to consider the case $M \le (n+2)/2$. Then on both phases we will be able to use orthogonal codes of length $n_1 = n/2$. The case of arbitrary M, such that $M = e^{o(n)}$, $n \to \infty$ can be considered replacing orthogonal codes by "almost" equidistant codes. Then all calculations remain essentially the same (see details in [1]).

Phase I. The transmitter uses the orthogonal code of M codewords $\{\boldsymbol{x}'_i\}$ of length n_1 such that $\|\boldsymbol{x}'_i\|^2 = A_1$.

Phase II. We set nonnegative numbers τ_2 and τ_3. Based on the received block $\boldsymbol{z}'$ and numbers τ_2, τ_3, the transmitter chooses $k = k(\boldsymbol{z}', \tau_2, \tau_3)$ most probable (for him) messages $k = 2, 3, 4$. Denote that set of messages as

$$\mathcal{S}^k = \left\{\boldsymbol{x}'^{(1)t}, \dots, \boldsymbol{x}'^{(k)t}\right\}, \qquad k = 2, 3, 4. \tag{10}$$

The code length n_1 for phase II we partition on two parts: of length 3 for selected $k \in \{2, 3, 4\}$ messages and of length $n_1 - 3$ for remaining $n - k$ messages, respectively. The transmitter uses the following code $\mathcal{C}'' = \mathcal{C}''(\boldsymbol{z}')$ with $\|\boldsymbol{x}''_j\|^2 = A_2$, $j = 1, \dots, M$.

1) If $d^{(3)t} - d^{(2)t} \geq 2A_1\tau_2$, then the transmitter selects two most probable (for him) messages θ_i, θ_j (i.e. $k = 2$) and uses for them opposite codewords $\boldsymbol{x}''_i = -\boldsymbol{x}''_j$ that have nonzero coordinates only at time instant $n_1 + 1$.

For remaining $M - 2$ messages $\{\theta_s\}$ the orthogonal code of $M - 2$ codewords $\{\boldsymbol{x}''_s\}$ of length $n_1 - 3$ is used. That code have zero components at time instants $n_1 + 1, n_1 + 2, n_1 + 3$, and all its codewords $\{\boldsymbol{x}''_s\}$ are orthogonal to the first two codewords $(\boldsymbol{x}''_i, \boldsymbol{x}''_j)$.

2) If $d^{(3)t} - d^{(2)t} < 2A_1\tau_2$, $d^{(4)t} - d^{(3)t} \geq 2A_1\tau_3$ then the transmitter selects three most probable (for him) messages $\theta_i, \theta_j, \theta_m$ (i.e. $k = 3$) and uses for them the 3-simplex code occupying time instants $n_1 + 1, n_1 + 2$.

For remaining $M - 3$ messages $\{\theta_s\}$ the orthogonal code of codewords $\{\boldsymbol{x}''_s\}$ of length $n_1 - 3$ is used. That code have zero components at time instants $n_1 + 1, n_1 + 2, n_1 + 3$, and all its codewords $\{\boldsymbol{x}''_s\}$ are orthogonal to the first three codewords $(\boldsymbol{x}''_i, \boldsymbol{x}''_j, \boldsymbol{x}''_m)$.

3) If $d^{(3)t} - d^{(2)t} < 2A_1\tau_2$, $d^{(4)t} - d^{(3)t} < 2A_1\tau_3$, then the transmitter selects four most probable (for him) messages $\theta_i, \theta_j, \theta_m, \theta_l$ (i.e. $k = 4$) and uses for them the 4-simplex code, occupying time instants $n_1 + 1, n_1 + 2, n_1 + 3$.

For remaining $M - 4$ messages $\{\theta_s\}$ the orthogonal code of codewords $\{\boldsymbol{x}''_s\}$ of length $n_1 - 3$ is used. That code have zero components at time instants $n_1 + 1, n_1 + 2, n_1 + 3$, and all its codewords $\{\boldsymbol{x}''_s\}$ are orthogonal to the first four codewords $(\boldsymbol{x}''_i, \boldsymbol{x}''_j, \boldsymbol{x}''_m, \boldsymbol{x}''_l)$.

This transmission method strengthens the method used in [1], [11]–[14], where only two or three messages were selected.

Note also that the set $\mathcal{S}^k$ of selected messages should be such that with high probability the true message $\theta_{\text{true}} \in \mathcal{S}^k$, but the number k is small as possible.

Remark 3. Introducing additional parameters $\tau_4, \ldots$, it is possible to strengthen the method used, but it gives not a big improvement of the results obtained. Much more improvement can be obtained using an increasing number of $N = N(\sigma)$ (see remark 2).

Decoding. Due to noise in the feedback channel the receiver does not know exactly codewords $\boldsymbol{x}'^{(1)t}, \boldsymbol{x}'^{(2)t}, \ldots$ and therefore it does not know the code used on phase II. But based on the received block $\boldsymbol{y}'$ it may evaluate probabilities of all possible codewords $\boldsymbol{x}'^{(1)t}, \boldsymbol{x}'^{(2)t}, \ldots$ and find the probabilities with which any code $\mathcal{C}''$ was used on phase II. It allows to the receiver, based on the full received block $\boldsymbol{y} = (\boldsymbol{y}', \boldsymbol{y}'')$, to find posterior probabilities $\{p(\boldsymbol{y}|\boldsymbol{x}_i)\}$ and make decision in favor of most probable message θ_i. Such full decoding is described in details in the next section.

§ 3. Full decoding and error probability P_{e}

Since $\|\boldsymbol{x}_i\|^2 = A$, $i = 1, \ldots, M$, for the likelihood ratio we have

$$\ln \frac{p\,(\boldsymbol{y}|\boldsymbol{x}_i)}{p\,(\boldsymbol{y}|\boldsymbol{x}_1)} = (\boldsymbol{x}_i - \boldsymbol{x}_1, \boldsymbol{y}).$$

If $\boldsymbol{x}_{\text{true}}$ is the true codeword then $\boldsymbol{y} = \boldsymbol{x}_{\text{true}} + \boldsymbol{\xi}$ and $\boldsymbol{\xi} = (\boldsymbol{\xi}', \boldsymbol{\xi}'') = (\xi_1, \ldots, \xi_n)$, where all $\{\xi_i\}$ are independent $\mathcal{N}(0, 1)$–Gaussian random variables. If $\boldsymbol{x}_{\text{true}} = \boldsymbol{x}_1$, then

$$\ln \frac{p\,(\boldsymbol{y}|\boldsymbol{x}_i)}{p\,(\boldsymbol{y}|\boldsymbol{x}_1)} = (\boldsymbol{x}_i - \boldsymbol{x}_1, \boldsymbol{\xi}) - \frac{1}{2}\|\boldsymbol{x}_i - \boldsymbol{x}_1\|^2$$

and

$$\ln \frac{p\left(\boldsymbol{y}|\boldsymbol{x}_3\right)}{p\left(\boldsymbol{y}|\boldsymbol{x}_2\right)} = (\boldsymbol{x}_3 - \boldsymbol{x}_2, \boldsymbol{\xi}) + (\boldsymbol{x}_3 - \boldsymbol{x}_2, \boldsymbol{x}_1),$$

where $(\boldsymbol{x}, \boldsymbol{\xi})$ is $\mathcal{N}(0, \|\boldsymbol{x}\|^2)$–Gaussian random variable.

For decoding error probability P_{e} we have

$$P_{\mathrm{e}} \le \frac{1}{M} \sum_{k=1}^{M} P_{\mathrm{e}k}, \tag{11}$$

where

$$P_{\mathrm{e}k} = \mathbf{P}\left\{ \max_{i \ne k} \ln \frac{p\left(\boldsymbol{y}\middle|\theta_i\right)}{p\left(\boldsymbol{y}\middle|\theta_k\right)} \ge 0 \middle| \theta_k \right\}, \quad k = 1, \ldots, M. \tag{12}$$

Denote $((\boldsymbol{x}'_i, \boldsymbol{x}'_1) = 0,\ i \ge 2)$

$$\begin{gathered} X_i = \ln \frac{p\left(\boldsymbol{y}'\middle|\theta_i\right)}{p\left(\boldsymbol{y}'\middle|\theta_1\right)} = (\boldsymbol{x}'_i - \boldsymbol{x}'_1, \boldsymbol{y}') = (\boldsymbol{x}'_i - \boldsymbol{x}'_1, \boldsymbol{\xi}') - A_1, \\ Y_i = \ln \frac{p\left(\boldsymbol{y}''\middle|\boldsymbol{y}', \theta_i\right)}{p\left(\boldsymbol{y}''\middle|\boldsymbol{y}', \theta_1\right)}. \end{gathered} \tag{13}$$

It is sufficient to investigate the value $P_{\mathrm{e}1}$, for which we have from (12)–(13)

$$\begin{gathered} P_{\mathrm{e}1} = \mathbf{P}\left\{ \max_{i \ge 2} (X_i + Y_i) \ge 0 \middle| \theta_1 \right\} \le \sum_{i \ge 2} \mathbf{P}\left\{ X_i + Y_i \ge 0 \middle| \theta_1 \right\} = \\ = \sum_{i \ge 2} \mathbf{E}_{\boldsymbol{y}'} \mathbf{P}\left\{ X_i + Y_i \ge 0 \middle| \boldsymbol{y}', \theta_1 \right\}. \end{gathered} \tag{14}$$

We can express the value Y_i via $\boldsymbol{y}'$ as follows. Since $\boldsymbol{x}''_i = \boldsymbol{x}''_i(\boldsymbol{z}')$ and $\boldsymbol{y}'' = \boldsymbol{x}''_1 + \boldsymbol{\xi}''$, then

$$\begin{gathered} e^{Y_i} = \frac{p\left(\boldsymbol{y}''\middle|\boldsymbol{y}', \theta_i\right)}{p\left(\boldsymbol{y}''\middle|\boldsymbol{y}', \theta_1\right)} = \mathbf{E}_{\boldsymbol{z}'|\boldsymbol{y}'} \frac{p\left(\boldsymbol{y}''\middle|\boldsymbol{z}', \boldsymbol{y}', \boldsymbol{x}''_i\right)}{p\left(\boldsymbol{y}''\middle|\boldsymbol{z}', \boldsymbol{y}', \boldsymbol{x}''_1\right)} = \\ = \mathbf{E}_{\boldsymbol{z}'|\boldsymbol{y}'} e^{(\boldsymbol{y}'', \boldsymbol{x}''_i - \boldsymbol{x}''_1)} = \mathbf{E}_{\boldsymbol{z}'|\boldsymbol{y}'} e^{(\boldsymbol{x}''_1, \boldsymbol{x}''_i - \boldsymbol{x}''_1) + (\boldsymbol{\xi}'', \boldsymbol{x}''_i - \boldsymbol{x}''_1)}, \end{gathered} \tag{15}$$

where the second equality is based on the fact that in both nominator and denominator the same code is used.

Remark 4. In order to apply the formula (15) it is necessary to know only the difference $\|\boldsymbol{x}''_i - \boldsymbol{x}''_1\|$ (depending on $\boldsymbol{z}'$). We do not need to know the whole code used on phase II. The selected group of messages of the code for phase II may consist of $2, 3, 4$ messages. For example, 3 messages are selected if 3 most probable messages are approximately equiprobable and all remaining messages are well separated from them (in metrics $d_i^{(t)}$).

We develop the right-hand side of the formula (15). The difference $\|\boldsymbol{x}''_i - \boldsymbol{x}''_1\|$ (depending on $\boldsymbol{z}'$) takes on one of 4 possible values (defined by partition groups, which those messages

belong to on phase II). It is convenient to separate those cases. Note that for all codewords of the k-simplex code we have

$$d_{ij} = \|\boldsymbol{x}_i'' - \boldsymbol{x}_j''\|^2 = 2A_2 k/(k-1), \qquad i \neq j.$$

Then denote

$$\begin{gathered} \delta_k = 2A_2 k/(k-1), \qquad k = 2, \ldots, K, \\ \delta_0 = 2A_2. \end{gathered} \tag{16}$$

In other words, δ_k, $k \geq 2$ is the codewords distance for k-simplex code, while d_0 is such distance for the orthogonal code. If $\boldsymbol{x}_1'', \boldsymbol{x}_i''$ belong to k-simplex code then

$$\begin{gathered} (\boldsymbol{x}_1'', \boldsymbol{x}_i'' - \boldsymbol{x}_1'') = -\delta_k/2 = -A_2 k/(k-1), \qquad k = 2, \ldots, K, \\ (\boldsymbol{x}_1'', \boldsymbol{x}_i'' - \boldsymbol{x}_1'') = -A_2 = -\delta_0/2, \qquad k = 0. \end{gathered}$$

The difference $\|\boldsymbol{x}_i'' - \boldsymbol{x}_1''\|$ may take on values $2A_2$ (corresponds to $k=0$) and δ_k, $k = 2, 3, 4$. Each value δ_k, $k = 2, 3, 4$ appears for phase II if a group of k messages was selected and both messages $\boldsymbol{x}_1', \boldsymbol{x}_i'$ belong to that group. In all other cases the value δ_0 is used.

Assuming $\theta_{\text{true}} = \theta_1$, introduce non-overlapping sets of random events

$$\mathcal{Z}_{i,k} = \left\{ \boldsymbol{z}' : \|\boldsymbol{x}_i'' - \boldsymbol{x}_1''\|^2 = \delta_k \right\}, \qquad k = 0, 2, 3, 4. \tag{17}$$

Denoting formally $\mathcal{Z}_{i,1} = \emptyset$, $i \geq 2$, we have $\{\boldsymbol{z}'\} = \sum_{k=0}^{4} \mathcal{Z}_{i,k}$ (here $\sum$ means the union of non-overlapping sets, and $\{\boldsymbol{z}'\}$ is the set of all possible outputs $\boldsymbol{z}'$).

We may continue (15) as follows

$$e^{Y_i} = \sum_{k=0}^{4} \mathbf{E}\left[e^{(\boldsymbol{x}_1'', \boldsymbol{x}_i'' - \boldsymbol{x}_1'') + (\boldsymbol{\xi}'', \boldsymbol{x}_i'' - \boldsymbol{x}_1'')}; \mathcal{Z}_{i,k} \middle| \boldsymbol{y}' \right] = \sum_{k=0}^{4} p_k e^{-\delta_k/2 + (\boldsymbol{\xi}'', \boldsymbol{x}_i'' - \boldsymbol{x}_1'')},$$

where $\mathbf{E}[\xi; \mathcal{A}] = \mathbf{E}(\xi \cdot I_{\{\mathcal{A}\}})$, $p_1 = 0$ and

$$p_k = p_k(\boldsymbol{y}') = p_k(\boldsymbol{\xi}') = \mathbf{P}\left(\mathcal{Z}_{i,k} \middle| \boldsymbol{y}' \right), \qquad k = 0, 2, 3, 4. \tag{18}$$

Then using (13) we have

$$e^{X_i + Y_i} = \sum_{k=0}^{4} p_k e^{-A_1 - \delta_k/2 + (\boldsymbol{x}_i' - \boldsymbol{x}_1', \boldsymbol{\xi}') + (\boldsymbol{\xi}'', \boldsymbol{x}_i'' - \boldsymbol{x}_1'')},$$

and therefore (since k takes on one of four possible values)

$$\begin{gathered} \mathbf{P}\left\{ X_i + Y_i \geq 0 \middle| \theta_1 \right\} = \mathbf{E}\mathbf{P}\left\{ e^{X_i + Y_i} \geq 1 \middle| \boldsymbol{y}', \theta_1 \right\} = \\ = \mathbf{E}\mathbf{P}\left\{ \sum_{k=0}^{4} p_k e^{-A_1 - \delta_k/2 + (\boldsymbol{x}_i' - \boldsymbol{x}_1', \boldsymbol{\xi}') + (\boldsymbol{x}_i'' - \boldsymbol{x}_1'', \boldsymbol{\xi}'')} \geq 1 \middle| \boldsymbol{y}', \theta_1 \right\} \leq \\ \leq \sum_{k=0}^{4} \mathbf{E}\mathbf{P}\left\{ \left[p_k e^{-A_1 - \delta_k/2 + (\boldsymbol{x}_i' - \boldsymbol{x}_1', \boldsymbol{\xi}') + (\boldsymbol{x}_i'' - \boldsymbol{x}_1'', \boldsymbol{\xi}'')} \geq 1/4 \right] \bigcap \mathcal{Z}_{i,k} \middle| \boldsymbol{y}', \theta_1 \right\} = \\ = \sum_{k=0}^{4} \mathbf{P}\left\{ [(\boldsymbol{x}_i' - \boldsymbol{x}_1', \boldsymbol{\xi}') + (\boldsymbol{x}_i'' - \boldsymbol{x}_1'', \boldsymbol{\xi}'') + \ln p_k(\boldsymbol{\xi}') \geq A_1 + \delta_k/2 - \ln 4] \bigcap \mathcal{Z}_{i,k} \middle| \theta_1 \right\}, \end{gathered} \tag{19}$$

where $\|\boldsymbol{x}_i'' - \boldsymbol{x}_1''\|^2 = \delta_k$ for the set $\mathcal{Z}_{i,k}$. Denote

$$(\boldsymbol{x}_i', \boldsymbol{\xi}') = \sqrt{A_1}\xi_i', \qquad (\boldsymbol{x}_i', \boldsymbol{\eta}') = \sqrt{A_1}\eta_i', \qquad i = 1, \ldots, M, \tag{20}$$

where all $\{\xi_i', \eta_i'\}$ are independent $\mathcal{N}(0,1)$-Gaussian random variables.

Since $(\boldsymbol{x}_i'' - \boldsymbol{x}_1'', \boldsymbol{\xi}'') \sim \sqrt{\delta_k}\xi''$ for the set $\mathcal{Z}_{i,k}$, we get from (19) and (20)

$$\begin{gathered} \mathbf{P}\left\{X_i + Y_i \geq 0 \middle| \theta_1\right\} \leq e^{o(1)} \sum_{k=0}^{4} P_{ik}, \\ P_{ik} = \mathbf{P}\left\{\sqrt{A_1}(\xi_i' - \xi_1') + \sqrt{d_k}\xi'' + \ln p_k(\boldsymbol{\xi}') \geq A_1 + \delta_k/2\right\}, \end{gathered} \tag{21}$$

where ξ'' does not depend on $\boldsymbol{\xi}'$ and $o(1) \to 0$ as $A_1 \to \infty$.

Probabilities $\{p_k(\boldsymbol{\xi}')\}$ and values P_{ik} from (21) are evaluated in the next section.

§ 4. Probabilities $p_k(\boldsymbol{\xi}')$ and values P_{ik}. Proof of Theorem

Let ξ be $\mathcal{N}(0,1)$–Gaussian random variable. We will regularly use simple inequality

$$\mathbf{P}(\xi \geq z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-u^2/2} du \leq e^{-z_+^2/2}, \qquad z \in \mathbb{R}^1, \tag{22}$$

and its natural generalization

L e m m a 1. 1) *Let* $(\xi_1, \ldots, \xi_K)$ *be independent* $\mathcal{N}(0,1)$*–Gaussian random variables and* $\mathcal{A} \subseteq \mathbb{R}^K$. *Then* ($\boldsymbol{x} = (x_1, \ldots, x_K)$, $\|\boldsymbol{x}\|^2 = x_1^2 + \ldots + x_K^2$)

$$\mathbf{P}\left((\xi_1, \ldots, \xi_K) \in \mathcal{A}\right) \leq \exp\left\{-\frac{1}{2} \inf_{\boldsymbol{x} \in \mathcal{A}} \|\boldsymbol{x}\|^2\right\}. \tag{23}$$

2) *Let* ξ, η *be* $\mathcal{N}(0,1)$*–Gaussian random variables and* $\mathbf{E}(\xi\eta) = \rho$. *Then*:

a) *if* $A - B\rho \geq 0$ and $B - A\rho \geq 0$ then

$$\mathbf{P}\left(\xi \geq A, \eta \geq B\right) \leq \mathbf{P}\left(\xi \geq \sqrt{\frac{A^2 + B^2 - 2AB\rho}{1 - \rho^2}}\right); \tag{24}$$

b) *otherwise*

$$\mathbf{P}\left(\xi \geq A, \eta \geq B\right) \leq \min\left\{\mathbf{P}(\xi \geq A), \mathbf{P}(\eta \geq B)\right\}. \tag{25}$$

P r o o f. 1) Let $\inf\limits_{\boldsymbol{x} \in \mathcal{A}} \|\boldsymbol{x}\| = r > 0$. Then $\mathcal{A} \subseteq \mathbb{R}^K \setminus S(r)$, where $S(r)$ – the ball of radius r. Therefore

$$\mathbf{P}\left((\xi_1, \ldots, \xi_K) \in \mathcal{A}\right) \leq \mathbf{P}\left\{(\xi_1, \ldots, \xi_K) \in \mathbb{R}^K \setminus S(r)\right\}.$$

Evaluating the last probability (using spherical coordinates) we get the formula (23).

2) We have

$$\mathbf{P}(\xi \ge A, \eta \ge B) \le \inf_{a \ge 0} \mathbf{P}(\xi + a\eta \ge A + aB) = \mathbf{P}\left(\xi \ge \frac{A + aB}{\sqrt{1 + a^2 + 2a\rho}}\right).$$

Minimizing the last expression over $a \ge 0$, we get the formulas (24)–(25). $\square$

Inequalities (23)–(25) give the exact logarithmic asymptotics in a natural asymptotic case.

In order to apply the formula (21), we consider sequentially the cases $k = 2, 0, 3, 4$.

1. Case $\mathbf{k = 2}$, $\delta_2 = 4A_2$. It is the simplest case and it takes place with probability close to 1. In that case $\boldsymbol{x}_1'', \boldsymbol{x}_i''$ compose the group $\mathcal{S}^2$ of two selected messages. Neglecting the term p_2 we get from (21)–(22)

$$\begin{gathered} P_{i2} \le \mathbf{P}\left\{(\boldsymbol{x}_i' - \boldsymbol{x}_1', \boldsymbol{\xi}') + 2\sqrt{A_2}\xi'' \ge A_1 + 2A_2 - \ln 3\right\} = \\ = \mathbf{P}\left\{\sqrt{2A_1 + 4A_2}\xi \ge A_1 + 2A_2 - \ln 3\right\} \le \exp\left\{-\frac{[A_1 + 2A_2 - \ln 3]_+^2}{4(A_1 + 2A_2)}\right\} \le \\ \le \sqrt{3}e^{-(A_1+2A_2)/4} = \sqrt{3}e^{-A_1(1+2\beta)/4}. \end{gathered} \tag{26}$$

Cases $k \ne 2$ are more computationally involved and in order to investigate them we will need the definition (10).

2. Case $\mathbf{k = 0}$, $\delta_0 = 2A_2$. It is the most computationally involved case. It takes place when the selected group of messages $\mathcal{S}^m$ contains not more than one of messages $\boldsymbol{x}_1'', \boldsymbol{x}_i''$. Then

$$P_{i0} = \sum_{m=2}^{4} P_{i0m}, \tag{27}$$

where $P_{i0m} = P\{k = 0, \mathcal{S}^m\}$, $m = 2, 3, 4$. We consider sequentially probabilities $\{P_{i0m}, m = 2, 3, 4\}$, starting with P_{i02}. Denote

$$d_{ij}' = \|\boldsymbol{x}_i' - \boldsymbol{x}_j'\|^2.$$

If $\theta_{\text{true}} = \theta_1$ then the formulas hold

$$\begin{gathered} d_i - d_j = d_{1i}' - d_{1j}' + 2(\boldsymbol{x}_j' - \boldsymbol{x}_i', \boldsymbol{\xi}'), \qquad i, j = 1, \ldots, M, \\ d_i - d_1 = d_{1i}' + 2(\boldsymbol{x}_1' - \boldsymbol{x}_i', \boldsymbol{\xi}'), \\ d_i^{(t)} - d_j^{(t)} = d_{1i}' - d_{1j}' + 2(\boldsymbol{x}_j' - \boldsymbol{x}_i', \boldsymbol{\xi}' + \sigma\boldsymbol{\eta}') = d_i - d_j + 2\sigma(\boldsymbol{x}_j' - \boldsymbol{x}_i', \boldsymbol{\eta}'), \\ d_i^{(t)} - d_1^{(t)} = d_{1i}' + 2\left(\boldsymbol{x}_1' - \boldsymbol{x}_i', \boldsymbol{\xi}' + \sigma\boldsymbol{\eta}'\right) = d_i - d_1 + 2\sigma(\boldsymbol{x}_1' - \boldsymbol{x}_i', \boldsymbol{\eta}'). \end{gathered} \tag{28}$$

If, in particular, $\boldsymbol{x}'^{(1)t} = \boldsymbol{x}_1'$, $\boldsymbol{x}'^{(2)t} = \boldsymbol{x}_2'$, and $\boldsymbol{x}'^{(3)t} = \boldsymbol{x}_i'$, $i \ge 3$, then in the case $\mathcal{S}^2$ it is necessary to have

$$d_3^{(t)} - d_2^{(t)} = 2(\boldsymbol{x}_2' - \boldsymbol{x}_i', \boldsymbol{\xi}' + \sigma\boldsymbol{\eta}') \ge 2A_1\tau_2. \tag{29}$$

In order to evaluate $p_0 = p_0(\boldsymbol{\xi}')$ from (18) notice that the main contribution to p_0 gives the case when the true message $\boldsymbol{x}'_1$ is selected, but the message $\boldsymbol{x}'_i$ is not. Moreover, in the case $\mathcal{S}^2$ maximum of p_0 is attained when $\boldsymbol{x}'^{(1)t} = \boldsymbol{x}'_1$, $\boldsymbol{x}'_i \notin \{\boldsymbol{x}'^{(1)t}, \boldsymbol{x}'^{(2)t}\}$. Taking into account symmetry of the orthogonal code $\{\boldsymbol{x}'_j\}$, we may assume that $\boldsymbol{x}'^{(2)t} = \boldsymbol{x}'_2$ and $\boldsymbol{x}'^{(3)t} = \boldsymbol{x}'_i$, $i \geq 3$. Since there are not more than M^3 variants of arranging messages $\boldsymbol{x}'_1, \boldsymbol{x}'_2, \boldsymbol{x}'_i$, then using (29), we have

$$\begin{gathered} p_0(\boldsymbol{\xi}') \leq M^3 \mathbf{P}\left((\boldsymbol{x}'_2 - \boldsymbol{x}'_i, \boldsymbol{\xi}' + \sigma\boldsymbol{\eta}') \geq A_1\tau_2 \Big| \{\xi'_i\}\right) \leq \\ \leq M^3 \mathbf{P}\left(\sigma(\boldsymbol{x}'_2 - \boldsymbol{x}'_i, \boldsymbol{\eta}') \geq A_1\tau_2 + \sqrt{A_1}\xi'_i - \sqrt{A_1}\xi'_2 \Big| \{\xi'_i\}\right) \leq \\ \leq M^3 \exp\left\{-\frac{(\sqrt{A_1}\tau_2 - \xi'_2 + \xi'_i)_+^2}{4\sigma^2}\right\}. \end{gathered} \tag{30}$$

Since $M = e^{o(A_1)}$, $A_1 \to \infty$ (see (8)–(9)), then from (21) and (30) for P_{i02} we get

$$\begin{gathered} P_{i02} \leq e^{o(A_1)} \mathbf{P}\left\{\sqrt{2A_2}\xi'' + \sqrt{A_1}(\xi'_i - \xi'_1) - \frac{(\sqrt{A_1}\tau_2 + \xi'_i - \xi'_2)_+^2}{4\sigma^2} \geq A_1 + A_2\right\} = \\ = e^{o(A_1)} \mathbf{P}\left\{\sqrt{A_1}\left(\sqrt{3+4\beta}\zeta + \zeta_2\right) - \frac{(\sqrt{A_1/2}\,\tau_2 + \zeta_2)_+^2}{\sigma^2\sqrt{2}} \geq A_1(1+\beta)\sqrt{2}\right\}, \end{gathered} \tag{31}$$

where we used the representations $\xi'_i - \xi'_2 = \sqrt{2}\zeta_2$, $\xi'_i - \xi'_1 = \zeta_2/\sqrt{2} + \sqrt{3/2}\,\xi$, $\xi \perp \zeta_2$ and $\sqrt{2A_2}\xi'' + \sqrt{3A_1/2}\,\xi = \sqrt{(3+4\beta)A_1/2}\zeta$, $\zeta \perp \zeta_2$ (here $\xi \perp \zeta$ means that Gaussian random variables ξ, ζ are orthogonal, i.e. independent).

Denoting $x\sqrt{A_1} = \zeta$, $y\sqrt{A_1} = \zeta_2$ and using the formula (23), we have from (31)

$$\begin{gathered} -2\ln P_{i02} \geq A_1 \inf_{(x,y)\in\mathcal{A}} \left(x^2 + y^2\right) + o(A_1), \\ \mathcal{A} = \left\{x, y : \sqrt{3+4\beta}x + y - \frac{(\tau_2/\sqrt{2} + y)_+^2}{\sigma^2\sqrt{2}} \geq (1+\beta)\sqrt{2}\right\}. \end{gathered} \tag{32}$$

Denoting $\mathcal{A}_1 = \left\{y : y \leq -\tau_2/\sqrt{2}\right\}$, first we have

$$\inf_{(x,y)\in(\mathcal{A}\cap\mathcal{A}_1)} (x^2+y^2) = \inf_{\substack{\sqrt{3+4\beta}x + y \geq (1+\beta)\sqrt{2} \\ y \leq -\tau_2/\sqrt{2}}} (x^2+y^2) = \inf_{\substack{\sqrt{3+4\beta}x + y \geq (1+\beta)\sqrt{2} \\ y = -\tau_2/\sqrt{2}}} (x^2+y^2),$$

i.e. infimum is attained on the border of $\mathcal{A}$. Therefore we may assume that $\tau_2/\sqrt{2} + y \geq 0$, omit the sign of positive part and replace (32) by

$$\begin{gathered} -2\ln P_{i02} \geq A_1 \inf_{(x,y)\in\mathcal{A}_2} \left(x^2 + y^2\right) + o(A_1), \\ \mathcal{A}_2 = \left\{x, y : x - \varepsilon(y+a)^2 \geq B\right\}, \end{gathered} \tag{33}$$

where

$$\varepsilon = \frac{1}{\sigma^2\sqrt{2(3+4\beta)}}, \qquad a = \frac{\sqrt{2}(\tau_2-\sigma^2)}{2}, \qquad B = \sqrt{\frac{2}{3+4\beta}}\left(1+\beta+\frac{2\tau_2-\sigma^2}{4}\right).$$

If $B \geq 0$, then for optimal x, y we need $x - \varepsilon(y+a)^2 = B$ and therefore (omitting $\varepsilon^2(y+a)^4$)

$$\inf_{(x,y)\in\mathcal{A}_2}\left(x^2+y^2\right) = \inf_y\left\{[\varepsilon(y+a)^2+B]^2+y^2\right\} \geq$$

$$\geq \inf_y\left\{2B\varepsilon(y+a)^2+B^2+y^2\right\} = B^2+a^2-\frac{a^2}{1+2B\varepsilon} \geq B^2+a^2-\frac{a^2}{2B\varepsilon}.$$

Therefore we get from (33) as $A_1 \to \infty$

$$-\ln P_{i02} \geq \frac{A_1}{2}\left\{\frac{(4+4\beta+2\tau_2-\sigma^2)^2}{8(3+4\beta)}+\frac{(\tau_2-\sigma^2)^2}{2}-\frac{(\tau_2-\sigma^2)^2(3+4\beta)\sigma^2}{4+4\beta+2\tau_2-\sigma^2}+o(1)\right\}. \quad (34)$$

We consider below only $\sigma^2 \leq 1$ and $\tau_2 \leq 4/9$. Then we can simplify the formula (34) as follows

$$\begin{aligned}-\ln P_{i02} &\geq \frac{A_1}{2}\left\{\frac{(2+2\beta+\tau_2)^2}{2(3+4\beta)}+\frac{\tau_2^2}{2}-\sigma^2\left[\frac{2+2\beta+\tau_2}{2(3+4\beta)}+\tau_2+\tau_2^2\right]+o(1)\right\} \geq \\ &\geq \frac{A_1}{4}\left[\frac{(2+2\beta+\tau_2)^2}{3+4\beta}+\tau_2^2+o(1)\right](1-\sigma^2) = \\ &= \frac{A_1(1+\beta)(1+\beta+\tau_2+\tau_2^2)(1-\sigma^2)}{3+4\beta}+o(A_1),\end{aligned} \quad (35)$$

since

$$\frac{(2+2\beta+\tau_2)^2}{2(3+4\beta)}+\frac{\tau_2^2}{2} \geq \frac{2+2\beta+\tau_2}{2(3+4\beta)}+\tau_2+\tau_2^2, \qquad \tau_2 \leq 4/9.$$

Consider the case of $\mathcal{S}^3$ and P_{i03}. Again, main contribution to p_0 and P_{i03} gives the case when the true message $\boldsymbol{x}'_1$ is selected, but $\boldsymbol{x}'_i$ is not. Maximum of p_0 and P_{i03} is attained when $\boldsymbol{x}'^{(1)t} = \boldsymbol{x}'_1, \boldsymbol{x}'^{(2)t} = \boldsymbol{x}'_2$, $\boldsymbol{x}'^{(3)t} = \boldsymbol{x}'_3$ and $i \geq 4$. Moreover, we need $d_1^{(t)} \leq d_2^{(t)} \leq d_3^{(t)} \leq d_i^{(t)}$ and $d_i^{(t)} \geq d_3^{(t)} + 2A_1\tau_3$. Then neglecting τ_2, similarly to (30)–(31) we have

$$\begin{aligned}p_0(\boldsymbol{\xi}') &\leq M^4\mathbf{P}\left((\boldsymbol{x}'_2-\boldsymbol{x}'_i, \boldsymbol{\xi}'+\sigma\boldsymbol{\eta}') \geq A_1\tau_3, (\boldsymbol{x}'_3-\boldsymbol{x}'_i, \boldsymbol{\xi}'+\sigma\boldsymbol{\eta}') \geq A_1\tau_3\Big|\{\xi'_i\}\right) \leq \\ &\leq M^4\mathbf{P}\left((\boldsymbol{x}'_2+\boldsymbol{x}'_3-2\boldsymbol{x}'_i, \boldsymbol{\xi}'+\sigma\boldsymbol{\eta}') \geq 2A_1\tau_3\Big|\{\xi'_i\}\right) = \\ &= M^4\mathbf{P}\left(\sigma(\boldsymbol{x}'_2+\boldsymbol{x}'_3-2\boldsymbol{x}'_i, \boldsymbol{\eta}') \geq 2A_1\tau_3+2\sqrt{A_1}\xi'_i-\sqrt{A_1}\xi'_2-\sqrt{A_1}\xi'_3\Big|\{\xi'_i\}\right) \leq \\ &\leq M^4\exp\left\{-\frac{(2\sqrt{A_1}\tau_3-\xi'_2-\xi'_3+2\xi'_i)_+^2}{6\sigma^2}\right\}\end{aligned}$$

and therefore (here ξ, ζ – independent $\mathcal{N}(0,1)$-Gaussian random variables)

$$P_{i03} \le e^{o(A_1)}\mathbf{P}\left\{\sqrt{A_1}(\xi_i' - \xi_1') + \sqrt{2A_2}\xi'' - \frac{(2\sqrt{A_1}\tau_3 - \xi_2' - \xi_3' + 2\xi_i')^2}{6\sigma^2} \ge A_1 + A_2\right\} \le$$

$$\le e^{o(A_1)}\mathbf{P}\left\{\sqrt{2A_1/3}\zeta + \sqrt{A_1/3}\zeta_1 - \sqrt{A_1}\xi_1' + \sqrt{2A_2}\xi'' - \frac{(2\sqrt{A_1}\tau_3 + \sqrt{6}\zeta)^2}{6\sigma^2} \ge A_1 + A_2\right\} =$$

$$= e^{o(A_1)}\mathbf{P}\left\{\sqrt{2A_1}\zeta + \sqrt{2(2A_1 + 3A_2)}\xi - \frac{\sqrt{3}(\sqrt{2A_1/3}\tau_3 + \zeta)_+^2}{\sigma^2} \ge \sqrt{3}(A_1 + A_2)\right\},$$

where we used the representations $2\xi_i' - \xi_2' - \xi_3' = \sqrt{6}\zeta$, $\xi_i' = \sqrt{2/3}\,\zeta + \zeta_1/\sqrt{3}$, $\zeta \perp \zeta_1$ and similar ones.

Denoting $y\sqrt{A_1} = \zeta$, $x\sqrt{A_1} = \xi$ and using the formula (23), similarly to (32) we have

$$-2\ln P_{i03} \ge A_1 \inf_{(x,y)\in\mathcal{A}} \left(x^2 + y^2\right) + o(A_1),$$

$$\mathcal{A} = \left\{x, y : x - \varepsilon(y + a)^2 \ge B\right\},$$

where we omitted the sign of the positive part (similarly to (33)) and where

$$\varepsilon = \frac{1}{\sigma^2}\sqrt{\frac{3}{2(2+3\beta)}}, \qquad a = \frac{2\tau_3 - \sigma^2}{\sqrt{6}}, \qquad B = \frac{6(1+\beta) + 4\tau_3 - \sigma^2}{2\sqrt{6(2+3\beta)}}.$$

Now similarly to the case P_{i02} we get as $A_1 \to \infty$

$$-\ln P_{i03} \ge$$

$$\ge \frac{A_1}{6}\left\{\frac{[6(1+\beta) + 4\tau_3 - \sigma^2]^2}{8(2+3\beta)} + \frac{(2\tau_3 - \sigma^2)^2}{2} - \frac{(2\tau_3 - \sigma^2)^2(2+3\beta)\sigma^2}{6(1+\beta) + 4\tau_3 - \sigma^2} + o(1)\right\}. \tag{36}$$

For $\sigma^2 \le 1$ the formula (36) can be simplified as follows

$$-\ln P_{i03} \ge \frac{A_1(1+\beta)(1-\sigma^2)}{4(2+3\beta)}[2 + 3\beta + (1 + 2\tau_3)^2]. \tag{37}$$

Consider P_{i04}. Maximum of p_0 and P_{i04} is attained when $\boldsymbol{x}'^{(j)t} = \boldsymbol{x}_j'$, $j = 1, \dots, 4$ and $i \ge 5$. Moreover, we need $d_1^{(t)} \le d_2^{(t)} \le d_3^{(t)} \le d_4^{(t)} \le d_i^{(t)}$. Then for any σ and $\beta \le 1/2$

$$\begin{aligned}
P_{i04} &\le e^{o(A_1)}\mathbf{P}\left\{\sqrt{A_1}(\xi_i' - \xi_1') + \sqrt{2A_2}\xi'' \ge A_1 + A_2, \xi_2' \ge \xi_i', \xi_3' \ge \xi_i', \xi_4' \ge \xi_i'\right\} \le \\
&\le e^{o(A_1)}\mathbf{P}\left\{\sqrt{A_1}\xi_i' + \sqrt{A_1 + 2A_2}\xi \ge A_1 + A_2, \xi_2 \ge \sqrt{3}\xi_i'\right\} \le \\
&\le e^{o(A_1)}\min_{a\ge 0}\mathbf{P}\left\{\sqrt{A_1}(1 - a\sqrt{3})\xi_i' + a\sqrt{A_1}\xi_2 + \sqrt{A_1 + 2A_2}\xi \ge A_1 + A_2\right\} = \\
&= e^{o(A_1)}\min_{a\ge 0}\mathbf{P}\left\{\sqrt{A_1[(1 - a\sqrt{3})^2 + a^2 + 1] + 2A_2}\xi \ge A_1 + A_2\right\} \le \\
&\le e^{o(A_1)}\mathbf{P}\left\{\sqrt{5A_1/4 + 2A_2}\xi \ge A_1 + A_2\right\} \le \\
&\le \exp\left\{-\frac{2(1+\beta)^2 A_1}{5 + 8\beta} + o(A_1)\right\} \le e^{-(1+\beta)A_1/3 + o(A_1)}.
\end{aligned} \tag{38}$$

Therefore for $\sigma^2 \le 1$ and $\tau_2 \le 4/9$ we get from (35), (37) and (38)

$$-\ln P_{i0} \ge \frac{(1+\beta)A_1}{4}\left[1+\min\left\{\frac{(1+2\tau_2)^2}{3+4\beta}, \frac{(1+2\tau_3)^2}{2+3\beta}, 1/3\right\}\right](1-\sigma^2)+o(A_1). \qquad (39)$$

Note that if τ_2, τ_3 satisfy conditions

$$\tau_2 \ge \frac{1}{\sqrt{15}+3} \approx 0.1455, \qquad \tau_3 \ge \frac{1}{2(\sqrt{42}+6)} \approx 0.04006, \qquad (40)$$

then for any $\beta \le 1/2$ the formula (39) takes the form

$$-\ln P_{i0} \ge \frac{A_1(1+\beta)(1-\sigma^2)}{3}+o(A_1). \qquad (41)$$

3. Case $\mathbf{k=3}$. $\delta_2 = 3A_2$. This case takes place if the group $\mathcal{S}^3$ of three messages was selected and $\boldsymbol{x}'_1, \boldsymbol{x}'_i \in \mathcal{S}^3$. Then $\|\boldsymbol{x}''_i - \boldsymbol{x}''_1\| = 3A_2$. Main contribution to $p_3(\boldsymbol{y}')$ and P_{i3} is given by case $\{\boldsymbol{x}'_1, \boldsymbol{x}'_i\} = \left\{\boldsymbol{x}'^{(1)t}, \boldsymbol{x}'^{(2)t}\right\}$. Moreover, since we are interested in the probability $\mathbf{P}\left\{X_i + Y_i \ge 0 \middle| \boldsymbol{y}', \theta_1\right\}$ and $\|\boldsymbol{x}''_i - \boldsymbol{x}''_1\|/A_2 = 3 > \|\boldsymbol{x}'_i - \boldsymbol{x}'_1\|/A_1 = 2$, then we may assume that $d_1^{(t)} \ge d_2^{(t)}$. More exactly, without loss of generality, we may assume that $i=2$ and $\boldsymbol{x}'_i = \boldsymbol{x}'_2 = \boldsymbol{x}'^{(1)t}$, $\boldsymbol{x}'_1 = \boldsymbol{x}'^{(2)t}$, $\boldsymbol{x}'_3 = \boldsymbol{x}'^{(3)t}$. Then first we have

$$P_{i3} \le M^3\mathbf{P}\{d_1 - d_2 \ge 0, d_3^{(t)} - d_1^{(t)} < 2A_1\tau_2 | \theta_1\} \le$$
$$\le M^3\mathbf{P}\{(\boldsymbol{x}'_2 - \boldsymbol{x}'_1, \boldsymbol{\xi}') + \sqrt{3A_2}\xi'' \ge A_1 + 3A_2/2, (\boldsymbol{x}'_3 - \boldsymbol{x}'_1, \boldsymbol{\xi}' + \sigma\boldsymbol{\eta}') \ge A_1(1-\tau_2)|\theta_1\} \le$$
$$= M^3\mathbf{P}\{\xi'_2 - \xi'_1 + \sqrt{3\beta}\xi'' \ge \sqrt{A_1}(1+3\beta/2), \xi'_3 - \xi'_1 - \sigma\sqrt{2}\eta \ge \sqrt{A_1}(1-\tau_2)\}.$$

Since $\xi'_2 - \xi'_1 + \sqrt{3\beta}\xi'' \sim \sqrt{2+3\beta}\,\xi$ and $\xi'_3 - \xi'_1 - \sigma\sqrt{2}\eta \sim \sqrt{2(1+\sigma^2)}\,\zeta$, where ξ, ζ – $\mathcal{N}(0,1)$-Gaussian random variables with $\mathbf{E}(\xi\zeta) = -1/\sqrt{2(2+3\beta)(1+\sigma^2)}$, then using the inequalities (24) and (22), we get as $A_1 \to \infty$

$$-\ln P_{i3} \ge -\ln \mathbf{P}\left\{\xi \ge \frac{1}{2}\sqrt{(2+3\beta)A_1}, \eta \ge (1-\tau_2)\sqrt{A_1/[2(1+\sigma^2)]}\right\} + o(A_1) \ge$$
$$\ge \frac{A_1}{4}\left[\frac{2+3\beta}{2} + \frac{(1-\tau_2)+(1-\tau_2)^2}{1+\sigma^2}\right] + o(A_1) \ge$$
$$\ge \frac{A_1(1+\beta)}{4}\left[1 + \frac{\beta}{2(1+\beta)} + \frac{2-3\tau_2}{(1+\beta)(1+\sigma^2)}\right] + o(A_1).$$

We limits ourselves only to values $\beta \le 1/2$, $\tau_2 \le 1/3$, $\sigma^2 \le 1$. Then

$$-\ln P_{i3} \ge \frac{A_1(1+\beta)}{3} + o(A_1). \qquad (42)$$

4. Case $\mathbf{k=4}$. $\delta_4 = 8A_2/3$. Similarly to $\mathcal{S}_3$ maximum of p_0 and P_{i4} is attained when $\boldsymbol{x}'^{(1)t} = \boldsymbol{x}'_i, \boldsymbol{x}'^{(2)t} = \boldsymbol{x}'_1$, $\boldsymbol{x}'^{(3)t} = \boldsymbol{x}'_3$, $\boldsymbol{x}'^{(4)t} = \boldsymbol{x}'_4$. Moreover, we need $d_i^{(t)} \le d_1^{(t)} \le d_3^{(t)} \le d_4^{(t)}$ and $d_3^{(t)} - d_1^{(t)} \le 2A_1\tau_2$, $d_4^{(t)} - d_3^{(t)} \le 2A_1\tau_3$. Then we have

$$P_{i4} \le M^4\mathbf{P}\{(\boldsymbol{x}'_2 - \boldsymbol{x}'_1, \boldsymbol{\xi}') + \sqrt{8A_2/3}\xi'' \ge A_1 + 4A_2/3,$$
$$(\boldsymbol{x}'_3 - \boldsymbol{x}'_1, \boldsymbol{\xi}' + \sigma\boldsymbol{\eta}') \ge A_1(1-\tau_2), (\boldsymbol{x}'_4 - \boldsymbol{x}'_3, \boldsymbol{\xi}' + \sigma\boldsymbol{\eta}') \ge -A_1\tau_3\} =$$
$$= M^4\mathbf{P}\Big\{\xi'_2 - \xi'_1 + \sqrt{8\beta/3}\xi'' \ge \sqrt{A_1}(1+4\beta/3),$$
$$\xi'_3 - \xi'_1 + \sigma(\eta'_3 - \eta'_1) \ge \sqrt{A_1}(1-\tau_2), \xi'_4 - \xi'_3 + \sigma(\eta'_4 - \eta'_3) \ge -\sqrt{A_1}\tau_3\Big\} =$$
$$= M^4\mathbf{P}\Big\{\sqrt{1+8\beta/3}\xi_2 - \xi'_1 \ge \sqrt{A_1}(1+4\beta/3),$$
$$\sqrt{1+\sigma^2}\xi_3 - \xi'_1 - \sigma\eta'_1 \ge \sqrt{A_1}(1-\tau_2), \sqrt{1+\sigma^2}(\xi_4 - \xi_3) \ge -\sqrt{A_1}\tau_3\Big\}$$

and therefore

$$-2\ln P_{i4} \ge A_1 \min_{\boldsymbol{z}\in\mathcal{A}} \|\boldsymbol{z}\|^2 + o(A_1), \qquad \boldsymbol{z} = (z_1, \dots, z_5),$$
$$\mathcal{A} = \Big\{\boldsymbol{z} : \sqrt{1+8\beta/3}z_2 - z_1 \ge 1 + 4\beta/3, \sqrt{1+\sigma^2}z_3 - z_1 - \sigma z_5 \ge 1 - \tau_2,$$
$$\sqrt{1+\sigma^2}(z_4 - z_3) \ge -\tau_3\Big\}.$$

Minimum is attained when there are equalities in all three inequalities. Then

$$z_4 = z_3 - \frac{\tau_3}{\sqrt{1+\sigma^2}}, \qquad \sigma z_5 = \sqrt{1+\sigma^2}z_3 - z_1 - 1 + \tau_2,$$

and after standard algebra we get

$$\min_{\boldsymbol{z}\in\mathcal{A}} \|\boldsymbol{z}\|^2 = \min_{z_1, y_3} \left\{ z_1^2 + \frac{(3z_1 + 3 + 4\beta)^2}{3(3+8\beta)} + \frac{y_3^2 + (y_3 - \tau_3)^2}{1+\sigma^2} + \frac{(y_3 - z_1 - 1 + \tau_2)^2}{\sigma^2} \right\} =$$
$$= (1-\tau_2)^2 + \frac{(3\tau_2 + 4\beta)^2}{3(3+8\beta)} + \frac{\sigma^2\tau_3^2}{(1+\sigma^2)(1+3\sigma^2)} - \frac{\left[1 - \tau_2 - \dfrac{3\tau_2 + 4\beta}{3+8\beta} - \dfrac{\tau_3}{1+3\sigma^2}\right]^2}{1 + \dfrac{3}{3+8\beta} + \dfrac{1}{\sigma^2} + \dfrac{1+\sigma^2}{\sigma^2(1+3\sigma^2)}} \ge$$
$$\ge (1-\tau_2)^2 + \frac{(3\tau_2 + 4\beta)^2}{3(3+8\beta)} - \sigma^2\left[1 - \tau_2 - \frac{3\tau_2 + 4\beta}{3+8\beta} - \frac{\tau_3}{1+3\sigma^2}\right]^2 \ge$$
$$\ge \left[(1-\tau_2)^2 + \frac{(3\tau_2 + 4\beta)^2}{3(3+8\beta)}\right](1-\sigma^2) = \frac{(3+4\beta)(3+4\beta-6\tau_2+6\tau_2^2)(1-\sigma^2)}{3(3+8\beta)},$$

since for $\tau_2 + \tau_3 \le 1$ we have

$$(1-\tau_2)^2 \ge \left[1 - \tau_2 - \frac{3\tau_2 + 4\beta}{3+8\beta} - \frac{\tau_3}{1+3\sigma^2}\right]^2.$$

Therefore if $\tau_2 + \tau_3 \le 1$, then

$$
\begin{aligned}
-\ln P_{i4} &\ge A_1 f_4(\beta, \tau_2)(1-\sigma^2) + o(A_1), \\
f_4(\beta, \tau_2) &= \frac{(3+4\beta)(3+4\beta-6\tau_2+6\tau_2^2)}{6(3+8\beta)}.
\end{aligned} \tag{43}
$$

Note that $f_4(1/2, \tau_2) \ge 1/2$, if $\tau_2 \le (15-\sqrt{105})/30 \approx 0.1584$.

Consider now the overall error probability P_i from (21). Assuming $\sigma^2 \le 1$, we set $\beta \le 1/2$. Then for τ_2, τ_3 satisfying conditions (40) and $\tau_2 \le 1/3$ we get from (41), (26), (42) and (43) as $A_1 \to \infty$

$$
\begin{aligned}
-\ln P_i &\ge \min\{-\ln P_{ik}, k = 0, 2, 3, 4\} + o(A_1) \ge \\
&\ge A_1(1-\sigma^2) \min\left\{\frac{(1+\beta)}{3}, \frac{1+2\beta}{4}, f_4(\beta, \tau_2)\right\} + o(A_1).
\end{aligned}
$$

We set $\beta = 1/2$. Then for any $\left(\sqrt{5/3}-1\right)/2 \approx 0.1455 \le \tau_2 \le (15-\sqrt{105})/30 \approx 0.1584$ and $\left(\sqrt{7/6}-1\right)/2 \approx 0.04006 \le \tau_3 \le 1-\tau_2$ we get as $A_1 \to \infty$ (i.e. as $n \to \infty$)

$$
-\ln P_i \ge \frac{A_1(1-\sigma^2)}{2} + o(n) = \frac{An(1-\sigma^2)}{3} + o(n). \tag{44}
$$

Since $P_{\mathrm{e}} \le P_{\mathrm{e}1}$ (see (11)) and $M = e^{o(n)}$, $n \to \infty$, then from (14), (21) and (44) we get

$$
-\ln P_{\mathrm{e}} \ge \frac{An(1-\sigma^2)}{3} + o(n),
$$

which completes the proof of Theorem. $\square$

REFERENCES

1. *Burnashev M. V., Yamamoto H.* On reliability function of Gaussian channel with noisy feedback: zero rate // Problems of Inform. Transm. 2012. V. 48, № 3. P. 3–22.

2. *Shannon C. E.* The Zero Error Capacity of a Noisy Channel // IRE Trans. Inform. Theory. 1956. V. 2. № 3. P. 8–19.

3. *Dobrushin R. L.* Asymptotic bounds on error probability for message transmission in a memoryless channel with feedback // Probl. Kibern. No. 8. M.: Fizmatgiz, 1962. P. 161–168.

4. *Horstein M.* Sequential Decoding Using Noiseless Feedback // IEEE Trans. Inform. Theory. 1963. V. 9. № 3. P. 136–143.

5. *Berlekamp E. R.*, Block Coding with Noiseless Feedback, Ph. D. Thesis, MIT, Dept. Electrical Enginering, 1964.

6. *Schalkwijk J. P. M., Kailath T.* A Coding Scheme for Additive Noise Channels with Feedback - I: No Bandwidth Constraint // IEEE Trans. Inform. Theory. 1966. V. 12. № 2. P. 172–182.

7. *Pinsker M. S.* The probability of error in block transmission in a memoryless Gaussian channel with feedback // Problems of Inform. Transm. 1968. V. 4. № 4. P. 3–19.

8. *Burnashev M. V.* Data transmission over a discrete channel with feedback: Random transmission time // Problems of Inform. Transm. 1976. V. 12. № 4. P. 10–30.

9. *Burnashev M. V.* On a Reliability Function of Binary Symmetric Channel with Feedback // Problems of Inform. Transm. 1988. V. 24. № 1. P. 3–10.

10. *Yamamoto H., Itoh R.* Asymptotic Performance of a Modified Schalkwijk–Barron Scheme for Channels with Noiseless Feedback // IEEE Trans. Inform. Theory. 1979. V. 25. № 6. P. 729–733.

11. *Burnashev M. V., Yamamoto H.* On BSC, Noisy Feedback and Three Messages // Proc. IEEE Int. Sympos. on Information Theory. Toronto, Canada. July, 2008. P. 886–889.

12. *Burnashev M. V., Yamamoto H.* On zero-rate error exponent for BSC with noisy feedback // Problems of Inform. Transm. 2008. V. 44. № 3. P. 33–49.

13. *Burnashev M. V., Yamamoto H.* Noisy Feedback Improves the BSC Reliability Function // Proc. IEEE Int. Sympos. on Information Theory. Seoul, Korea. June–July, 2009. P. 886–889.

14. *Burnashev M. V., Yamamoto H.* On reliability function of BSC with noisy feedback // Problems of Inform. Transm. 2010. V. 46. № 2. P. 2–23.

15. *Xiang Y., Kim Y.-H.* On the AWGN channel with noisy feedback and peak energy constraint // Proc. IEEE International Symposium on Information Theory. Austin, Texas, June 2010. P. 256-259.

16. *Shannon C. E.* Probability of Error for Optimal Codes in Gaussian Channel // Bell System Techn. J. 1959. V. 38. № 3. P. 611–656.

Burnashev Marat Valievich
Kharkevich Institute for Information Transmission Problems,
Russian Academy of Sciences, Moscow
burn@iitp.ru
Yamamoto Hirosuke
School of Frontier Sciences
The University of Tokyo, Japan
hirosuke@ieee.org